

Stochastic equations for a self-regulating gene

Martin Jansen¹

March 14, 2014

Abstract

Expression of cellular genes is regulated by binding of transcription factors to their promoter, either activating or inhibiting transcription of a gene. Particularly interesting is the case when the expressed protein regulates its own transcription. In this paper the features of this self-regulating process are investigated. In the here presented model the gene can be in two states. Either a protein is bound to its promoter or not. The steady state distributions of protein during and at the end of both states are analyzed. Moreover a powerful numerical method based on the corresponding master equation to compute the protein distribution in the steady state is presented and compared to an already existing method. Additionally the special case of self-regulation, in which protein can only be produced, if one of these proteins is bound to the promoter region, is analyzed. Furthermore a self-regulating gene is compared to a similar gene, which also has two states and produces the same amount of proteins but is not regulated by its protein-product.

Keywords and phrases: Gene expression, chemical reaction network, activated gene, gene regulatory networks, master equation

AMS Subject Classification: 92C42 60J28 65C40

1 Introduction

Gene expression denotes the cascade of certain reactions in a cell, causing the synthesis of RNA and proteins originated from a gene. Important steps in this cascade of reactions are transcription i.e. production of RNA from the corresponding gene by RNA polymerase and translation i.e. production of protein from RNA through ribosomes. The usage of mathematical models for analysis of gene expression is a broad and well explored field (Ay and Arnosti, 2011). The regulation of gene expression was first described by Jacob and Monod (1961), who studied the regulation of lactose metabolism by the lac operon in *Escherichia Coli*. A proper regulation of gene expression is an essential feature for a cell. For example unicellular organisms need to react rapidly on a change of their environment and multicellular organisms need to control the differentiation of cells. In a common form of gene regulation proteins (so called regulatory proteins) cause a change of the binding-rate of DNA polymerase to

¹Universitätsklinik Freiburg, Albert-Ludwigs University of Freiburg, Hugstetter Str. 55, D-79106 Freiburg, Germany, e-mail: martin.jansen@uniklinik-freiburg.de

the promoter region, leading to either activation or inhibition of transcription of the gene (Ptashne, 1992). In this paper regulatory proteins, directly associated with the product of the corresponding gene expression are considered. So a self-regulation through a feedback takes places.

There are many different approaches describing genetic regulatory systems such as a self-regulating gene mathematically. These approaches use for example techniques like Boolean networks, ordinary and partial differential equations and directed graphs (De Jong, 2002). In this context a gene regulated by its own protein-product is a relative simple system.

Stochasticity plays an important role in gene regulation (McAdams and Arkin, 1999; Elowitz et al., 2002; Paulsson, 2005). Especially the promoter fluctuation, therefore the stochastic activation and deactivation of a gene contributes to the cellular noise (Zhu and Salahub, 2008). This is caused amongst others by the low number of gene copies, which equals in diploid organisms in general two, and the fact that the gene is situated at the beginning of the reaction cascade. So proteins easily inherit stochastic effects. Additionally there are stochastic effects caused by the low number of RNA copies.

A common approach to model gene expression is to make the same assumptions as in a chemical reaction network (Gunawardena, 2003). So each reaction underlies mass-action kinetics, which simplifies the analysis a lot. Hence it is assumed that the protein production steps are exponentially distributed, with rates depending on the state of the gene.

In the here presented model the self-regulating gene has two states: either the promoter region of the gene is bound or unbound by a regulatory protein. Protein number and the state of the gene are the only quantities considered, so transcription and translation are lumped together as one reaction. The model assumptions are the same as used by Hornos et al. (2005). Peccoud and Ycart (1995) considered a similar and simpler model, in which switching of the gene was independent of its product, so no feedback takes places.

We are interested in analysing and computing the steady state distribution of protein numbers during and at the end of the bound and the unbound state.

Hornos et al. (2005) introduced a method to compute the protein distributions during both states. As key result another method to compute this distribution is presented and compared with the approach of Hornos.

Outline: After describing the model, the distribution of the system in the steady state is considered. Linear dependencies and recursions for this distribution (Theorem 1), and a recursive description of its moments in the unbound and the bound state are derived (Theorem 2). These recursions can be used to compute the correct protein distribution in the steady state (Theorem 3). Next the distributions of the proteins at the beginning and the end of the bound state are considered (Theorems 4-6). Hereafter a kind of extreme positive feedback is considered, where no protein is produced during the unbound state, which always leads to an extinction of the proteins after a finite number of unbound states. The number and duration of the bound-unbound cycles until the protein is extinct, given n proteins at time 0 and other features, is investigated (Theorem 7). In Section 2.2 an algorithm based on the recursions of Theorem 1, which computes the protein distribution effectively, is presented and discussed. In Section 2.3 the influence of feedback on gene expression is analysed by comparing a model without feedback with a model with feedback. A discussion is

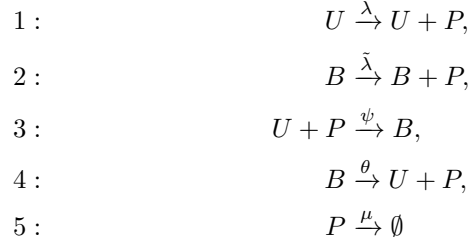
stated in Section 3, in which this method is compared to the one used by Hornos et al. (2005). Furthermore the impact of stochasticity to the model is illustrated using an example. Proofs are given in Section 4.

2 Main results

In this section a model for a self regulating gene is described and analysed. An algorithm to compute the steady state distribution of the system is deduced.

2.1 Model and results

The model refers to the expression of a single gene. The set of chemical reactions



is considered for reaction rates $\lambda, \psi, \theta, \mu > 0$ and $\tilde{\lambda} \geq 0$. Here, P is a protein which is expressed by a gene, which comes in two states, U and B . Precisely, U and B denote that the gene is unbound or bound by one protein molecule P . If the gene is in state U , the protein is expressed at rate λ , while it is expressed at rate $\tilde{\lambda}$, if the gene is in state B . Clearly, the case $\lambda > \tilde{\lambda}$ ($\lambda < \tilde{\lambda}$) is a negative (positive) feedback loop, where a protein inhibits (enhances) its own transcription.

Here the gene produces directly the protein, so the dynamics of mRNA and other intermediate steps in gene expression are not modeled.

Let N_t be the total number of proteins (either free or bound to the gene) at time t , and $A_t = 0$ or $A_t = 1$ if the promoter is bound to the protein or not. By the law of mass action, the process $(X_t)_{t \geq 0}$ with $X_t = (N_t, A_t)$ obeys the time-change equations

$$\begin{aligned}
 N_t &= N_0 + Y_1 \left(\int_0^t \lambda 1_{A_s=1} ds \right) + Y_2 \left(\int_0^t \tilde{\lambda} 1_{A_s=0} ds \right) - Y_3 \left(\int_0^t \mu N_s (1 - 1_{A_s=0}) ds \right), \\
 A_t &= A_0 - Y_4 \left(\int_0^t \psi 1_{A_s=1} N_s ds \right) + Y_5 \left(\int_0^t \theta 1_{A_s=0} ds \right)
 \end{aligned}$$

for independent, unit rate Poisson processes Y_1, \dots, Y_5 . The fact that this system of equations has a unique steady state solution $\lim_{t \rightarrow \infty} X_t = X$ with $(N, A) := X$ follows from Markov chain theory (Anderson and Kurtz, 2011).

It holds for all $t \geq 0$, that $P\{X_t = (0, 0)\} = 0$, since in the bound state one

protein is always bound to the gene. The master equation for $n \in \mathbb{N}$ is

$$\begin{aligned}\frac{\partial P\{X_t = (n, 1)\}}{\partial t} &= P\{X_t = (n-1, 1)\}\lambda - P\{X_t = (n, 1)\}(n(\psi + \mu) + \lambda) \\ &\quad + P\{X_t = (n+1, 1)\}(n+1)\mu + P\{X_t = (n, 0)\}\theta, \\ \frac{\partial P\{X_t = (n, 0)\}}{\partial t} &= P\{X_t = (n-1, 0)\}\tilde{\lambda} - P\{X_t = (n, 0)\}(\theta + (n-1)\mu + \tilde{\lambda}) \\ &\quad + P\{X_t = (n+1, 0)\}n\mu + P\{X_t = (n, 1)\}n\psi.\end{aligned}\tag{2.1}$$

It is straightforward to derive linear relationships between the states in the steady state.

Theorem 1 (Recursions and equations for the steady state protein distribution). *Consider the steady state distribution X and let $n \in \mathbb{N}$. The decay of protein, when there are n proteins, equals the production of protein, when there are $n-1$ proteins*

$$\begin{aligned}P\{X = (n, 1)\}\mu n + P\{X = (n, 0)\}\mu(n-1) \\ = P\{X = (n-1, 1)\}\lambda + P\{X = (n-1, 0)\}\tilde{\lambda}.\end{aligned}\tag{2.2}$$

The gene switches as often in the state $A = 1$, as it switches in the state $A = 0$

$$\theta \sum_{i=0}^{\infty} P\{X = (i, 0)\} = \psi \sum_{i=0}^{\infty} i P\{X = (i, 1)\}.\tag{2.3}$$

A state $(n, 1)$ is left as often, as it is entered

$$\begin{aligned}P\{X = (n, 1)\}(n(\psi + \mu) + \lambda) \\ = P\{X = (n-1, 1)\}\lambda + P\{X = (n+1, 1)\}(n+1)\mu + P\{X = (n, 0)\}\theta,\end{aligned}\tag{2.4}$$

and a state $(n, 0)$ is left as often, as it is entered

$$\begin{aligned}P\{X = (n, 0)\}((n-1)\mu + \theta + \tilde{\lambda}) \\ = P\{X = (n-1, 0)\}\tilde{\lambda} + P\{X = (n+1, 0)\}n\mu + P\{X = (n, 1)\}\psi n.\end{aligned}\tag{2.5}$$

With these recursions the first moment of the proteins can be characterised and recursions for the bound and the unbound state can be given:

Corollary 2.1 (First moments and recursions for the states $A = 1$ and $A = 0$). *The expectation values of N , $1_{A=1}N$ and $1_{A=0}N$ are*

$$\mathbb{E}[N] = P\{A = 0\}\left(1 + \frac{\tilde{\lambda}}{\mu}\right) + P\{A = 1\}\frac{\lambda}{\mu},\tag{2.6}$$

$$\mathbb{E}[1_{A=1}N] = P\{A = 0\}\frac{\theta}{\psi},\tag{2.7}$$

$$\mathbb{E}[1_{A=0}N] = P\{A = 0\}\left(1 + \frac{\tilde{\lambda}}{\mu} - \frac{\theta}{\psi}\right) + P\{A = 1\}\frac{\lambda}{\mu}.\tag{2.8}$$

Furthermore it holds in the unbound state

$$P\{X = (0, 1)\} = \frac{\mu}{\lambda} P\{X = (1, 1)\}. \quad (2.9)$$

It holds that

$$\frac{\min(\lambda, \tilde{\lambda} + \mu)}{\mu} \leq \mathbb{E}[N] \leq \frac{\max(\lambda, \tilde{\lambda} + \mu)}{\mu}. \quad (2.10)$$

For $A = 1$ and $n \geq 1$ the recursion

$$\begin{aligned} P\{X = (n+2, 1)\} = & \\ \frac{1}{n(n+2)\mu^2} \Big[& P\{X = (n+1, 1)\} \mu \left(n((n+1)(\psi + \mu) + \lambda) + (\tilde{\lambda} + \theta)(n+1) \right) \\ & - P\{X = (n, 1)\} \left(\tilde{\lambda}(n(\psi + \mu) + \lambda) + n\lambda\mu + \lambda\theta \right) + P\{X = (n-1, 1)\} \lambda\tilde{\lambda} \Big] \end{aligned} \quad (2.11)$$

holds. For $A = 0$ and $n \geq 1$ the recursion

$$\begin{aligned} P\{X = (n+2, 0)\} = & \\ \frac{1}{n(n+1)\mu^2} \Big[& -P\{X = (n, 0)\} \left(n\tilde{\lambda}(\mu + \psi) + \lambda((n-1)\mu + \theta + \tilde{\lambda}) \right) \\ & + nP\{X = (n+1, 0)\} \mu \left(n(\psi + \mu) + \theta + \lambda + \tilde{\lambda} \right) + P\{X = (n-1, 0)\} \lambda\tilde{\lambda} \Big] \end{aligned} \quad (2.12)$$

holds.

It is possible to compute higher moments of N conditioned on both states in a recursive way.

Theorem 2 (Recursions for higher moments). *Let $s \in \mathbb{N}$, the higher moments of N and $1_{A=1}N$ are given by*

$$\begin{aligned} \mathbb{E}[1_{A=1}N^{s+1}] = & \\ \frac{1}{\psi} \Big(& \theta \mathbb{E}[1_{A=0}N^s] + \sum_{j=1}^s \binom{s}{j} ((-1)^j \mu \mathbb{E}[1_{A=1}N^{s-j+1}] + \lambda \mathbb{E}[1_{A=1}N^{s-j}]) \Big), \end{aligned} \quad (2.13)$$

$$\begin{aligned} \mathbb{E}[N^{s+1}] = & \\ \frac{\lambda}{\mu} \mathbb{E}[N^s 1_{A=1}] + \frac{\mu + \tilde{\lambda}}{\mu} \mathbb{E}[N^s 1_{A=0}] - \sum_{j=1}^s \binom{s}{j} (-1)^j (\mathbb{E}[N^{s-j+1}] - \mathbb{E}[N^{s-j} 1_{A=0}]). \end{aligned} \quad (2.14)$$

These moments exist for all $s \in \mathbb{N}$, as in both states N is dominated by the Poisson distribution with parameter $\max(\lambda, \tilde{\lambda} + \mu)/\mu$.

Corollary 2.2 (The second moment of the protein number distribution). *The second moments of the protein distribution in the steady state are characterised by*

$$\mathbb{E}[1_{A=1}N^2] = \frac{1}{\psi} \left(P\{A = 0\} \theta \left(1 + \frac{\tilde{\lambda}}{\mu} - \frac{\theta}{\psi} - \frac{\mu}{\psi} \right) + \lambda P\{A = 1\} \left(1 + \frac{\theta}{\mu} \right) \right),$$

$$\mathbb{E}[N^2] = P\{A=0\} \left(\frac{\theta(\lambda - \tilde{\lambda})}{\psi\mu} + 1 + \frac{\tilde{\lambda}}{\mu} \left(3 + \frac{\tilde{\lambda}}{\mu} \right) \right) + \frac{\lambda}{\mu} P\{A=1\} \left(2 + \frac{\tilde{\lambda}}{\mu} \right).$$

Theorem 2 and the Corollaries 2.1 and 2.2 are derived from the master equation (2.1) and provide explicit formulas for moments of X linear dependent on $P\{A=1\}$. However an explicit formula for $P\{A=1\}$ is not given. The following Theorem 3 can be used to compute $P\{A=1\}$ and more general the distribution X numerically.

Theorem 3 (Relation between X and the recursion (2.11)). *For $n \geq 1$ set*

$$x_{n+2} = \frac{1}{n(n+2)\mu^2} \left[x_{n+1}\mu \left(n((n+1)(\psi + \mu) + \lambda) + (\tilde{\lambda} + \theta)(n+1) \right) - x_n \left(\tilde{\lambda}(n(\psi + \mu) + \lambda) + n\lambda\mu + \lambda\theta \right) + x_{n-1}\lambda\tilde{\lambda} \right]. \quad (2.15)$$

For each combination of $\psi, \theta, \mu, \lambda \in \mathbb{R}^{>0}, \tilde{\lambda} \in \mathbb{R}^{\geq 0}, n \geq 1$ and $x_{n-1} > 0$ (respectively $x_{n+1} > 0$), there is only one combination of $x_n, x_{n+1} > 0$ (respectively $x_n, x_{n-1} > 0$), so that the recursively determined sequence defined by (2.15) determined by the values x_{n-1}, x_n and x_{n+1} at position $n-1, n$ and $n+1$ respectively has the features of a measure.

Due to the linearity of (2.15) (respectively (2.11), which is (2.15) evaluated with the probabilities of the unbound state) in x_{n+1}, x_n and x_{n-1} this sequence is proportional to the sequence $(P\{X=(n,1)\})_{n \in \mathbb{N}}$. It is straightforward to compute a sequence proportional (with the same factor) to $(P\{X=(n,0)\})_{n \in \mathbb{N}}$ with the equation (2.4). Given both sequences the probability distribution of X can be obtained by normalisation.

Furthermore using Theorem 3 the interval, in which $P\{A=1\}$ lies can be estimated.

Corollary 2.3 (Estimation of $P\{A=1\}$). *For $\tilde{\lambda} = \lambda$ the probability distributions of the protein number in both states are Poisson and it holds*

$$P\{A=1\} = \frac{\mu\theta}{\mu\theta + \lambda\psi}. \quad (2.16)$$

Generally it holds

$$\frac{\mu\theta}{\mu\theta + \max(\lambda, \tilde{\lambda})\psi} \leq P\{A=1\} \leq \frac{\mu\theta}{\mu\theta + \min(\lambda, \tilde{\lambda})\psi}. \quad (2.17)$$

Using Theorem 2 more estimations can be made, as for all $s \in \mathbb{N}$ the terms $\mathbb{E}[N^s], \mathbb{E}[N^s 1_{A=1}]$ and $\mathbb{E}[N^s 1_{A=0}]$ can be expressed as linear terms in $P\{A=1\}$. So claims like $\mathbb{E}[N^{s+1} 1_{A=0}] \geq \mathbb{E}[N^s 1_{A=0}]$ lead to estimations for $P\{A=1\}$. However these estimations get complexer with increasing s and are in general not better than (2.17).

For example considering $\mathbb{E}[N 1_{A=0}] \geq P\{A=0\}$ it can be derived by using (2.7) for $\lambda\psi + \theta\mu \geq \tilde{\lambda}\psi$, that

$$\frac{\mu\theta - \tilde{\lambda}\psi}{\psi(\lambda - \tilde{\lambda}) + \mu\theta} \leq P\{A=1\}.$$

For $\tilde{\lambda} = 0$ this corresponds to (2.17). Only if additionally $\tilde{\lambda} > \lambda$ and $\tilde{\lambda}^2 - \tilde{\lambda}\frac{\mu\theta}{\psi} + \mu\lambda < 0$, this estimation is better than (2.17).

The previous part provides a characterisation of X . Theorem 3 provides the basis for an algorithm presented in Section 2.2, which can compute the probability distribution of X . Next the distributions of proteins at the end of the unbound and bound state are considered.

Definition 2.4. *Let C_1 be the protein number at the end of the bound state and B_1 the protein number at the end of the unbound state.*

Let C_0 (respectively B_0) be the protein number at the beginning of a unbound (respectively bound) state, which ends with B_1 (respectively C_1) proteins. Let C and B be the corresponding steady state distributions.

It is clear, that $P\{C_1 = 0\} = P\{B_1 = 0\} = 0$.

Theorem 4 (the bound state). *The expected number of proteins at the end of a bound state, starting with $b \geq 1$ proteins is given by*

$$\mathbb{E}[C_1|B_0 = b] = \frac{b\theta + \tilde{\lambda} + \mu}{\theta + \mu}. \quad (2.18)$$

There is a linear dependence between the expected number of C and B in the equilibrium

$$\mathbb{E}[C] = \frac{\mathbb{E}[B]\theta + \tilde{\lambda} + \mu}{\theta + \mu}. \quad (2.19)$$

For $c, b \geq 1$ the conditioned probability distribution of C_1 given $B_0 = b$ is

$$P\{C_1 = c|B_0 = b\} = \sum_{n=\max(0, b-c)}^{b-1} \binom{b-1}{n} \frac{\theta \left(\frac{\tilde{\lambda}}{\mu}\right)^{c-b+n}}{(c-b+n)!} \int_0^\infty e^{-t(\theta+n\mu) - \frac{\tilde{\lambda}}{\mu}(1-e^{-t\mu})} (1-e^{-t\mu})^{c-1} dt.$$

If the case $\tilde{\lambda} = 0$, where there is no production in the bound state, is considered, $P\{C_1 = c|B_0 = b\}$ can be displayed explicitly for $c \geq 1$

$$P\{C_1 = c|B_0 = b\} = \theta \binom{b-1}{c-1} \sum_{s=0}^{c-1} \binom{c-1}{s} \frac{(-1)^s}{\theta + \mu(b-1-s)}.$$

Considering (2.19) it can be seen, that if λ is changed while all other parameters are fixed, $\mathbb{E}[C]$ and $\mathbb{E}[B]$ are linearly dependent.

Theorem 5 (The distribution C). *The distribution of proteins at the end of the bound state in the steady state equals the distribution of proteins during the bound state.*

Theorem 6 (The distribution B). *The first two moments of B in the steady*

state are

$$\begin{aligned}\mathbb{E}[B] &= \frac{\mathbb{E}[C](\theta + \mu) - \tilde{\lambda} - \mu}{\theta}, \\ \mathbb{E}[B^2] &= \frac{\mathbb{E}[C^2](\theta + 2\mu)}{\theta} - \frac{\mu\mathbb{E}[B]}{\theta + \mu} \left(3\mu\theta + \frac{2\tilde{\lambda}}{\mu} \right) - \left(2 + \frac{2\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) \\ &\quad + \frac{\theta + 2\mu}{\theta + \mu} \left(3 + \frac{5\tilde{\lambda}}{\mu} + 2 \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) - (\theta + 2\mu)\theta \left(1 + 3\frac{\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right).\end{aligned}$$

If $\lambda = \tilde{\lambda}$, it follows directly with Corollary 2.3, that $\mathbb{E}[C] = \mathbb{E}[B] = 1 + \frac{\lambda}{\mu}$.

Remark 2.5 (computation of B). Using Markov chain theory it is possible but numerically costly to compute $P\{B_1 = b|C_0 = c\}$ (or $P\{C_1 = c|B_0 = b\}$). For this purpose the transition probabilities of getting from $C_0 = c$ to $B_1 = b$ in $m \in \mathbb{N}$ reaction steps have to be computed. Hence the distribution of B in the steady state can be computed exactly with some numerical effort given the distribution $P\{X = (i, 0)\}$ for $i \in \mathbb{N}$, as by Theorem 5

$$P\{B = b\} = \sum_{c=0}^{\infty} P\{B_1 = b|C_0 = c\} P\{X = (c, 0)\}.$$

Next the case $\lambda = 0$ is considered. Here it is certain, that during some unbound state all protein decay. Hence the steady state only consists of the absorbing state with no protein, so $P\{X = (0, 1)\} = 1$.

Definition 2.6. Let $\lambda = 0$. Consider X_t starting at $t = 0$. Let S be the number of bound-unbound cycles until extinction and T_i be the length of time of i bound-unbound-cycles.

Theorem 7 (the case $\lambda = 0$). Let $\lambda = 0$ and $i \geq 2$, then conditioned on the protein number $n \geq 1$ at time 0, the probability distribution of S is given by

$$P\{S = 1|X_0 = (n, 0)\} = \sum_{m=1}^{\infty} P\{C_1 = m|B_0 = n\} \left(\frac{\mu}{\mu + \psi} \right)^m,$$

$$\begin{aligned}P\{S = i|X_0 = (n, 0)\} &= \\ \sum_{m=1}^{\infty} \frac{\psi P\{C_1 = m|B_0 = n\}}{\mu + \psi} \sum_{j=1}^m \left(\frac{\mu}{\mu + \psi} \right)^{m-j} P\{S = i-1|X_0 = (j, 0)\}.\end{aligned}$$

Furthermore it holds

$$\begin{aligned}\mathbb{E}[T_1|X_0 = (n, 0), S > 1] &= \\ \frac{1}{\theta} + \sum_{m=1}^{\infty} \frac{\psi P\{C_1 = m|B_0 = n\}}{\mu + \psi} \sum_{j=1}^m \left(\frac{\mu}{\mu + \psi} \right)^{m-j} \left[\frac{1}{\mu} \sum_{i=j+1}^m \frac{1}{i} + \frac{1}{\psi j} \right], \\ \mathbb{E}[T_i|X_0 = (n, 0), S > i] &= \frac{1}{\theta} + \sum_{m=1}^{\infty} \frac{\psi P\{C_1 = m|B_0 = n\}}{\mu + \psi} \sum_{j=1}^m \left(\frac{\mu}{\mu + \psi} \right)^{m-j} \\ &\quad \left(\frac{1}{\mu} \sum_{k=j+1}^m \frac{1}{k} + \frac{1}{\psi j} + \mathbb{E}[T_{i-1}|X_0 = (j, 0), S > i-1] \right).\end{aligned}$$

For the first two moments of the number of bound-unbound cycles until extinction starting with $n \geq 1$ the following fixed-point equations hold:

$$\begin{aligned}\mathbb{E}[S|X_0 = (n, 0)] &= \sum_{m=1}^{\infty} \frac{\psi P\{C_1 = m|B_0 = n\}}{\mu + \psi} \sum_{j=1}^m \left(\frac{\mu}{\mu + \psi}\right)^{m-j} (\mathbb{E}[S|X_0 = (j, 0)] + 1), \\ \mathbb{E}[S^2|X_0 = (n, 0)] &= 2\mathbb{E}[S|X_0 = (n, 0)] \\ &+ \sum_{m=1}^{\infty} P\{C_1 = m|B_0 = n\} \sum_{j=1}^m \left(\frac{\mu}{\mu + \psi}\right)^{m-j} \frac{\psi (\mathbb{E}[S^2|X_0 = (j, 0)] - 1)}{\mu + \psi}.\end{aligned}$$

If the unbound state starting with n proteins ends after a finite time, it holds:

$$\mathbb{E}[B_1|C_0 = n] = \frac{\mu + \psi}{\psi} \left(1 - \left(\frac{\mu}{\mu + \psi}\right)^n \left(1 + \frac{n\psi}{\mu + \psi}\right)\right).$$

Let $i, n \in \mathbb{N}$. For the computation of $P\{S = i|X_0 = (n, 0)\}$, $\mathbb{E}[S|X_0 = (n, 0)]$, $\mathbb{E}[S^2|X_0 = (n, 0)]$ and $\mathbb{E}[T_i|X_0 = (n, 0), i < S]$ the conditional probabilities given in Remark 2.5 are needed. Hence their numerical computation is costly.

Corollary 2.7. *Let $\lambda = 0$. Let C_1 be the distributions of proteins at the beginning of a unbound state and C_2 be the corresponding distribution at the beginning of the following unbound state. Given the first unbound state ends after a finite time and starts with m proteins, it holds for $m \geq 1$*

$$\begin{aligned}\mathbb{E}[C_2|C_1 = m] &= \frac{\tilde{\lambda} + \mu}{\theta + \mu} - \frac{\theta}{\theta + \mu} \frac{\mu + \psi}{\psi} \left(\left(\frac{\mu}{\mu + \psi}\right)^m \left(m \frac{\psi}{\mu + \psi} + 1\right) - 1 \right), \\ \frac{\tilde{\lambda} + \mu}{\theta + \mu} + \frac{\theta}{\theta + \mu} \frac{\psi}{\mu + \psi} &\leq \mathbb{E}[C_2|C_1 = m] \leq \frac{\tilde{\lambda} + \mu}{\theta + \mu} + \frac{\theta}{\theta + \mu} \frac{\mu + \psi}{\psi}.\end{aligned}$$

Given i bound-unbound cycles, the distribution of totally produced proteins can be calculated:

$$P\{m \text{ produced proteins} | i \text{ cycles}\} = \left(\frac{\theta}{\theta + \tilde{\lambda}}\right)^i \left(\frac{\tilde{\lambda}}{\theta + \tilde{\lambda}}\right)^m \binom{m+i-1}{i}.$$

2.2 Algorithm for the computation of the distribution X

Next a method to compute the equilibrium distribution of protein numbers in the bound and unbound state is presented.

Definition 2.8 (the recursion R). *For $i, n \in \mathbb{N}$ and $x, y, z \in \mathbb{R}$ let $R_{i,n}(x, y, z)$ be the recursion (2.15) determined by fixed x, y, z at position $n - 1, n, n + 1$ evaluated at position i .*

Hence given $n \in \mathbb{N}$ it holds

$$R_{n-1,n}(x, y, z) = x, R_{n,n}(x, y, z) = y, R_{n+1,n}(x, y, z) = z,$$

for all $i \geq n$

$$\begin{aligned} R_{i+2,n}(x, y, z) &= \frac{1}{i(i+2)\mu^2} \left[-R_{i,n}(x, y, z) \left(\tilde{\lambda}(i(\psi + \mu) + \lambda) + i\lambda\mu + \lambda\theta \right) \right. \\ &\quad \left. + R_{i+1,n}(x, y, z)\mu \left(i((i+1)(\psi + \mu) + \lambda) + (\tilde{\lambda} + \theta)(i+1) \right) + R_{i-1,n}(x, y, z)\lambda\tilde{\lambda} \right] \end{aligned} \quad (2.20)$$

and for all $1 \leq i \leq n-1$ (and $\tilde{\lambda} \neq 0$)

$$\begin{aligned} R_{i-1,n}(x, y, z) &= \frac{1}{\lambda\tilde{\lambda}} \left[-R_{i+1,n}(x, y, z)\mu \left(i((i+1)(\psi + \mu) + \lambda) + (\tilde{\lambda} + \theta)(i+1) \right) \right. \\ &\quad \left. + i(i+2)\mu^2 R_{i+2,n}(x, y, z) + R_{i,n}(x, y, z) \left(\tilde{\lambda}(i(\psi + \mu) + \lambda) + i\lambda\mu + \lambda\theta \right) \right]. \end{aligned} \quad (2.21)$$

The goal is to find $\hat{x}, \hat{y}, \hat{z}$, so that $P\{X = (i, 1)\} = R_{i,n}(\hat{x}, \hat{y}, \hat{z})$ holds for all $i \in \mathbb{N}$. In the following the cases $n = 1$ and $n \geq 2$ are distinguished. By Theorem 3 the solution is unique. Due to the linearity of (2.15) it holds for $x \neq 0$, that

$$R_{i,n}(x, y, z) = x R_{i,n}\left(1, \frac{y}{x}, \frac{z}{x}\right). \quad (2.22)$$

So if the starting value at position $n-1$ is set to one, by Theorem 3 there are unique (\tilde{y}, \tilde{z}) , so that the sequence $(R_{i,n}(1, \tilde{y}, \tilde{z}))_{i \in \mathbb{N}}$ has the features of a measure.

Let Y be Poisson distributed with parameter $\max\left(\frac{\lambda}{\mu}, \frac{\tilde{\lambda}}{\mu} + 1\right)$. It is easy to see that there exists a $m \in \mathbb{N}$, so that $P\{X = (n, 1)\} \leq P\{Y = n\}$ for all $n \geq m$. Hence for n big enough it holds $P\{X = (n, 1)\} \approx 0$.

2.2.1 starting the algorithm at position zero

If the recursion is started at the beginning, choose $\tilde{x} = 1$ and by (2.9) $\tilde{y} = \frac{\lambda}{\mu}$. So only the unique \tilde{z} , for which the sequence $(R_{i,1}\left(1, \frac{\lambda}{\mu}, \tilde{z}\right))_{i \in \mathbb{N}}$ has the features of a measure, needs to be determined. It holds $\tilde{z} = \frac{P\{X=(2,1)\}}{P\{X=(0,1)\}}$ by (2.22). Furthermore due to the linearity of the recursion there are $q_i, r_i \in \mathbb{R}$ for each $i \in \mathbb{N}$, with $R_{i,1}\left(1, \frac{\lambda}{\mu}, z\right) = q_i z + r_i$. Given two unequal points z_1, z_2 , it is straightforward to compute for $i \in \mathbb{N}$

$$\begin{aligned} q_i &= \frac{R_{i,1}\left(1, \frac{\lambda}{\mu}, z_1\right) - R_{i,1}\left(1, \frac{\lambda}{\mu}, z_2\right)}{z_1 - z_2}, \\ r_i &= R_{i,1}\left(1, \frac{\lambda}{\mu}, z_1\right) - q_i z_1. \end{aligned} \quad (2.23)$$

Let m be large enough (as just described), it can be expected that the unique \tilde{z} fulfills $R_{m,1}\left(1, \frac{\lambda}{\mu}, \tilde{z}\right) = q_m \tilde{z} + r_m \approx 0$. Hence given two unequal points z_1, z_2 , the following approximation z' for \tilde{z} holds

$$z' = -\frac{r_m}{q_m} = z_1 - R_{m,1}\left(1, \frac{\lambda}{\mu}, z_1\right) \frac{z_1 - z_2}{R_{m,1}\left(1, \frac{\lambda}{\mu}, z_1\right) - R_{m,1}\left(1, \frac{\lambda}{\mu}, z_2\right)}. \quad (2.24)$$

Due to numerical inaccuracy it is advisable to compute z' more than once with different starting values z_1 and z_2 . Considering the recursion, it can be shown that for all $z \neq \tilde{z}$ the sequence $\left(|R_{i,1}(1, \frac{\lambda}{\mu}, z)|\right)_{i \in \mathbb{N}}$ is unbounded.

The corresponding sequence of the bound state can be computed using (2.2) and (2.5). To compute the corresponding probabilities, both sequences are scaled with their total sum.

Now the algorithm can be outlined for $n = 1$:

1. Choose $m \gg n$ so that $\frac{\max(\lambda, \tilde{\lambda} + \mu)^m}{\mu^m m!} e^{-\frac{\max(\lambda, \tilde{\lambda} + \mu)}{\mu}} \approx 0$
2. Choose $z_1, z_2 > 0$ with $z_1 \neq z_2$
3. Compute q_m, r_m using (2.23)
4. Compute z' using (2.24)
5. Compute the sequence $\left(R_{i,1}\left(1, \frac{\lambda}{\mu}, z'\right)\right)_{0 \leq i \leq m}$ using the recursion (2.20)
6. Compute the bound state-sequence $(S_i)_{0 \leq i \leq m}$ using the recursion (2.5), (2.12) and the sequence $\left(R_{i,1}\left(1, \frac{\lambda}{\mu}, z'\right)\right)_{0 \leq i \leq m}$
7. Set for all $0 \leq i \leq m$

$$P\{X = (i, 1)\} = \frac{R_{i,1}\left(1, \frac{\lambda}{\mu}, z'\right)}{\sum_{j=0}^m R_{j,1}\left(1, \frac{\lambda}{\mu}, z'\right) + S_j},$$

$$P\{X = (i, 0)\} = \frac{S_i}{\sum_{j=0}^m R_{j,1}\left(1, \frac{\lambda}{\mu}, z'\right) + S_j}.$$

Computation of the protein distribution can lead to numerical problems depending on the parameters $\psi, \lambda, \tilde{\lambda}, \mu$ and θ . If $\frac{\lambda}{\mu}$ is large, the probabilities $P\{X = (1, 1)\}$ and $P\{X = (2, 1)\}$ may lie near the machine precision. In such cases it may be beneficial to either start the recursion at a position $n \geq 2$, or if $\tilde{\lambda}$ is relatively small to use an analogous algorithm based on the recursion for the bound state.

2.2.2 starting the algorithm in $n \geq 2$

In some cases it is beneficial to start the recursion at a position $n \geq 2$ and to compute the $P\{X = (i, 1)\}$ for all positions $i \in \mathbb{N}$, with $i < n - 1$ and $i > n + 1$. This approach works only for $\tilde{\lambda} > 0$. In contrast to the case where the recursion is started at the beginning, two instead of one parameters have to be determined. However again the linearity of the recursion can be used. Analogously to the previous case a $m \gg n$ is chosen, for which the corresponding probability $P\{X = (m, 1)\}$ is expected to be nearly zero. Set the parameter $\tilde{x} = 1$. Additionally condition (2.9) should hold for the backward recursion, hence \tilde{y}, \tilde{z} are searched, who fulfill the conditions

$$\begin{aligned} \lambda R_{0,n}(1, \tilde{y}, \tilde{z}) &= \mu R_{1,n}(1, \tilde{y}, \tilde{z}), \\ R_{m,n}(1, \tilde{y}, \tilde{z}) &\approx 0. \end{aligned} \tag{2.25}$$

Consider a fixed $n \in \mathbb{N}$. For each $i \in \mathbb{N}$, there are $a_i, b_i, c_i \in \mathbb{R}$ with $R_{i,n}(1, y, z) = a_i y + b_i z + c_i$. Given y_1, y_2, z_1, z_2 with $y_1 \neq y_2$ and $z_1 \neq z_2$ it is straightforward to compute for $i \in \mathbb{N}$

$$\begin{aligned} a_i &= \frac{R_{i,n}(1, y_1, z_1) - R_{i,n}(1, y_2, z_1)}{y_1 - y_2}, \\ b_i &= \frac{R_{i,n}(1, y_1, z_1) - R_{i,n}(1, y_1, z_2)}{z_1 - z_2}, \\ c_i &= -a_i y_1 - b_i z_1 + R_{i,n}(1, y_1, z_1). \end{aligned} \quad (2.26)$$

By using the conditions (2.25) an approximation (y', z') for (\tilde{y}, \tilde{z})

$$\begin{aligned} z' &= \frac{a_m \tilde{c} - c_m \tilde{a}}{b_m \tilde{a} - a_m \tilde{b}}, \\ y' &= -\frac{c_m}{a_m} - \frac{b_m}{a_m} \frac{a_m \tilde{c} - c_m \tilde{a}}{b_m \tilde{a} - a_m \tilde{b}} \end{aligned} \quad (2.27)$$

is obtained with

$$\tilde{a} := \frac{\lambda}{\mu} a_0 - a_1, \tilde{b} := \frac{\lambda}{\mu} b_0 - b_1 \text{ and } \tilde{c} := \frac{\lambda}{\mu} c_0 - c_1. \quad (2.28)$$

Analogously we need to compute the distribution of the bound state and to normalize the solution. However due to numerical reasons, the choice of y_1, y_2, z_1, z_2 and n is important and should be considered.

Now the algorithm for $n \geq 2$ and $\tilde{\lambda} > 0$ can be outlined:

1. Choose $m \gg n$ so that $\frac{\max(\lambda, \tilde{\lambda} + \mu)^m}{\mu^m m!} e^{-\frac{\max(\lambda, \tilde{\lambda} + \mu)}{\mu}} \approx 0$
2. Choose $y_1, z_1, y_2, z_2 > 0$ with $y_1 \neq y_2$ and $z_1 \neq z_2$
3. Compute $a_m, a_1, a_0, b_m, b_1, b_0, c_m, c_1, c_0$ using (2.26) and $\tilde{a}, \tilde{b}, \tilde{c}$ using (2.28)
4. Compute y', z' using (2.27)
5. Compute the sequence $(R_{i,n}(1, y', z'))_{0 \leq i \leq m}$ using the recursions (2.20) and (2.21)
6. Compute the bound state-sequence $(S_i)_{0 \leq i \leq m}$ using the recursion (2.5), (2.12) and the sequence $(R_{i,n}(1, y', z'))_{0 \leq i \leq m}$
7. Set for all $0 \leq i \leq m$

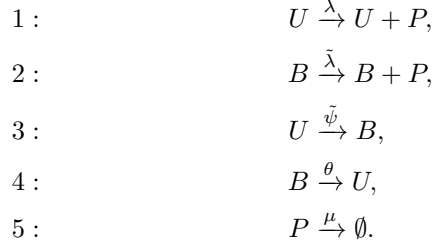
$$\begin{aligned} P\{X = (i, 1)\} &= \frac{R_{i,n}(1, y', z')}{\sum_{j=0}^m R_{j,n}(1, y', z') + S_j}, \\ P\{X = (i, 0)\} &= \frac{S_i}{\sum_{j=0}^m R_{j,n}(1, y', z') + S_j}. \end{aligned}$$

2.3 The impact of the feedback on the system

In this section the impact of feedback on the system is analysed. Therefore a non-feedback model (NFM) describing gene expression is introduced. To compare this model with a model describing feedback, the feedback-model described in Section 2.1 is modified slightly. This feedback-model is called modified feedback model (MFM).

2.3.1 the non-feedback model (NFM)

The NFM equals the gene expression model except for two points. First all protein can decay during the bound state and second the unbound states length is exponentially distributed with a parameter $\tilde{\psi}$. Thus its duration does not depend on the current number of proteins. The NFM reactions are



Hence in contrast to the model described in Section 2.1, only first order reactions are considered, which simplifies the model alot. Peccoud and Ycart (1995) introduced a special case of this model, where $\tilde{\lambda} = 0$. Analogously to the feedback-model let M and \tilde{A} be the protein number and the state of the gene in the NFM in the steady state. It holds $P\{\tilde{A} = 1\} = \frac{\theta}{\theta + \tilde{\psi}}$. Using the equation corresponding to (2.3) and the methods of Theorem 2 it can be derived, that

$$\mathbb{E}[M] = \frac{\lambda}{\mu} P\{\tilde{A} = 1\} + \frac{\tilde{\lambda}}{\mu} P\{\tilde{A} = 0\}, \quad (2.29)$$

$$\mathbb{E}[M^2] = \frac{\lambda}{\mu} \left(P\{\tilde{A} = 1\} + \mathbb{E}[1_{\tilde{A}=1} M] \right) + \frac{\tilde{\lambda}}{\mu} \left(P\{\tilde{A} = 0\} + \mathbb{E}[1_{\tilde{A}=0} M] \right). \quad (2.30)$$

In contrast to the feedback-model all moments can be derived exactly. In the following the expected number of proteins in the unbound state is derived. Advancing similar to Theorem 4 and using the notation B and C of Definition 2.4 for the protein number in the steady state at the end of the unbound and bound state in the NFM, it can be proved, that

$$\mathbb{E}[B] = \frac{\mathbb{E}[C] \tilde{\psi} + \lambda}{\mu + \tilde{\psi}}, \quad (2.31)$$

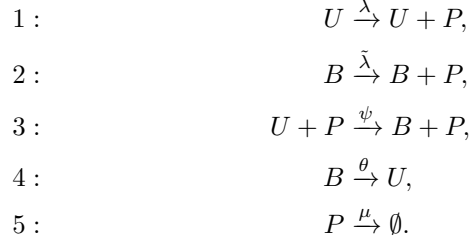
$$\mathbb{E}[C] = \frac{\mathbb{E}[B] \theta + \tilde{\lambda}}{\mu + \theta}. \quad (2.32)$$

Furthermore both states durations are exponentially distributed. So as seen in Theorem 5 it holds, $\mathbb{E}[M|\tilde{A} = 1] = \mathbb{E}[B]$ and $\mathbb{E}[M|\tilde{A} = 0] = \mathbb{E}[C]$. Hence it can be derived, that

$$\mathbb{E}[M|\tilde{A} = 1] = \frac{\lambda(\theta + \mu) + \tilde{\psi}\tilde{\lambda}}{\mu(\mu + \tilde{\psi} + \theta)}. \quad (2.33)$$

2.3.2 the modified feedback model (MFM)

To compare the NFM to a corresponding and suitable feedback-model, the MFM is constructed. Therefore the model described in Section 2.1 is changed in the following way: the protein bound to the gene can decay, which does not interrupt the bound states duration. The reactions are



Let \bar{N} and \bar{A} be the protein number and the state of the gene in the steady state. Analogously to the methods used in this paper (compare to Corollary 2.1 and Theorem 2) it can be derived, that

$$\mathbb{E} [\bar{N} | \bar{A} = 1] = \frac{P \{\bar{A} = 0\} \theta}{P \{\bar{A} = 1\} \tilde{\psi}}, \quad (2.34)$$

$$\mathbb{E} [\bar{N}] = \frac{\lambda}{\mu} P \{\bar{A} = 1\} + \frac{\tilde{\lambda}}{\mu} P \{\bar{A} = 0\}, \quad (2.35)$$

$$\mathbb{E} [\bar{N}^2] = \frac{\lambda}{\mu} (P \{\bar{A} = 1\} + \mathbb{E} [1_{\bar{A}=1} \bar{N}]) + \frac{\tilde{\lambda}}{\mu} (P \{\bar{A} = 0\} + \mathbb{E} [1_{\bar{A}=0} \bar{N}]). \quad (2.36)$$

2.3.3 Comparism of both models

To compare M and \bar{N} , fix λ, μ, θ and ψ and set

$$\tilde{\psi} = \theta \frac{P \{\bar{A} = 0\}}{P \{\bar{A} = 1\}}. \quad (2.37)$$

As consequence it follows $P \{\tilde{A} = 1\} = P \{\bar{A} = 1\}$ and $\mathbb{E} [M] = \mathbb{E} [\bar{N}]$, so in the steady state both models have the same protein in- and efflux.

Proposition 2.9 (Relation between the variance and the expectation conditional on the state of the gene between the MFM and the NFM). *If (2.37) holds, it follows that*

$$\mathbb{E} [\bar{N} | \bar{A} = 1] < \mathbb{E} [M | \tilde{A} = 1], \quad (2.38)$$

$$\begin{aligned}
 V(\bar{N}) &\leq V(M), \text{ if } \lambda \geq \tilde{\lambda} \text{ and} \\
 V(\bar{N}) &\geq V(M), \text{ if } \lambda \leq \tilde{\lambda}.
 \end{aligned} \quad (2.39)$$

2.4 Impact of stochasticity on the model

Hornos et al. (2005) discuss the impact of stochasticity on the system of the self-regulating gene, by comparing the features of the model to the corresponding features of the common deterministic mass-action approach (Ackers et al., 1982), which neglects the promoter fluctuation, hence the discrete distinction between the bound and unbound state. Here the relevance of stochasticity is illustrated by estimating the maximal difference between the model and the deterministic mass-action approach, if $\tilde{\lambda} = 0$ is fixed.

Corresponding to N in the stochastic case, let \tilde{N} be the distribution of protein-number in the steady state in the deterministic approach. In the steady state protein-production equals decay of protein, so it holds

$$0 = \frac{\theta}{\theta + \mathbb{E}[\tilde{N}]\psi} \left(\lambda - \mu \mathbb{E}[\tilde{N}] \right) + \frac{\mathbb{E}[\tilde{N}]\psi}{\theta + \mathbb{E}[\tilde{N}]\psi} \left(\tilde{\lambda} - \mu \left(\mathbb{E}[\tilde{N}] - 1 \right) \right).$$

Thus it holds that

$$\mathbb{E}[\tilde{N}] = \frac{1}{2\mu\psi} \left(\tilde{\lambda}\psi + \mu(\psi - \theta) + \sqrt{\left(\tilde{\lambda}\psi + \mu(\psi - \theta) \right)^2 + 4\lambda\theta\mu\psi} \right).$$

Proposition 2.10 (Difference of the expected steady state protein number between the deterministic and the stochastic approach). *Let $c := \frac{\lambda}{\mu}$. Consider $\tilde{\lambda} = 0$, and fixed θ, ψ, c . It holds*

$$\max_{\mu \in \mathbb{R}^{>0}} \left| \mathbb{E}[N] - \mathbb{E}[\tilde{N}] \right| \geq \frac{1}{2\psi} \left(\psi - \theta + \sqrt{(\psi - \theta)^2 + 4\theta\psi c} \right) - c \frac{\theta + \psi}{\theta + c\psi}.$$

$\mathbb{E}[N]$ can be smaller or greater than $\mathbb{E}[\tilde{N}]$.

3 Discussion

In this work the same model as described by Hornos et al. (2005) is considered. However additional to the analysis of the distribution of proteins in the unbound and bound state, the distributions at the end of these states and the case $\lambda = 0$ are analysed.

Furthermore a different method to calculate the protein distribution during the two states is presented. It uses only a recursive description of the probability distribution derived from the master equation.

In this discussion the results are outlined and numerical examples are given. Next the here presented method to compute the protein-distribution is compared to the method of Hornos. Finally the impact of stochasticity on the model is discussed.

In Theorem 2 it is shown, that all higher moments of the bound and unbound state equal linear combinations of the expectation values of lower moments in the bound and unbound state. Hence given the correct $P\{A = 1\}$ all moments can be calculated. However we have no formula for $P\{A = 1\}$ given the reaction rates. In Section 2.2 an algorithm to compute X and therewith $P\{A = 1\}$ is presented.

It is shown, that the steady state protein distribution at the end of the bound state equals the distribution during the bound state. However the steady state protein distribution at the end of the unbound state is more complex to compute. Here the time in the unbound state depends on the evolution of protein numbers during the state. Only the first two moments are given in this paper. Higher moments are linearly dependent on moments of equal and lower order of C . Figure 1 displays the distribution of proteins at the end and during both states. In the bound state they are equal by Theorem 5. In the unbound state the protein-mass is higher at the end of the state than during the state. The probability to have no proteins at the end of the state is zero.

Setting $\lambda = 0$ leads to the fact, that the protein number in the steady state is

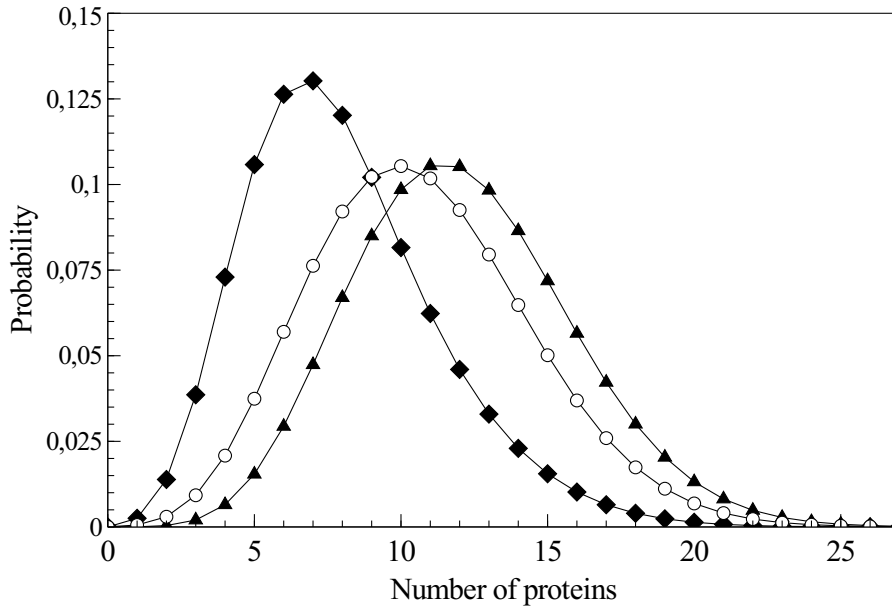


Figure 1: Distribution of proteins at the beginning (diamonds), during (circles) and at the end of the unbound state (triangles) in the steady state. A positive feedback is considered. Parameter values are $\lambda = 15, \tilde{\lambda} = 5, \mu = 1, \theta = 0.5, \psi = 0.1$. The distribution of proteins during the bound state equals the distribution at the beginning of the unbound state.

0, as there is always the chance in the unbound state, that each protein decays, so no protein can initiate the bound state, in which proteins can be produced with rate $\tilde{\lambda}$. This special case is not common in gene regulation. However it gives some indication of how the system behaves for very small λ . The time and number of bound and unbound states, until the last protein decays, dependent on the initial number of proteins at time 0 is considered. In Theorem 7 and Corollary 2.7 equations, which describe implicitly and explicitly certain parameters of this special positive feedback model, are derived. The probability $P\{C_1 = m | B_0 = n\}$ appearing in Theorem 7 can be numerically approached with Theorem 4.

Figure 2 displays the probabilities of protein extinction after a corresponding number of bound states, depending on the number of proteins at time zero.

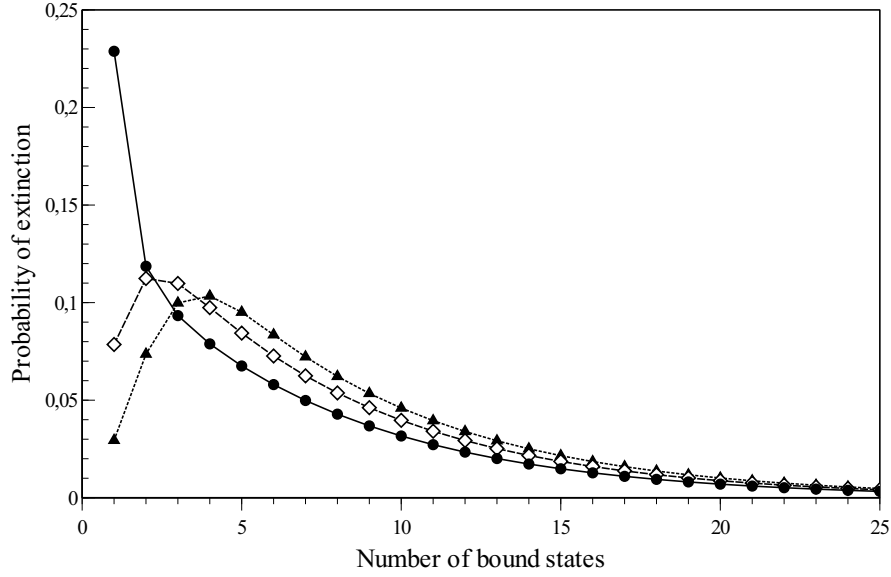


Figure 2: Probability distribution of number of bound-unbound cycles until extinction of protein, given $n = 1$ (solid line, circles), 16 (dashed line, diamonds), 64 (dotted line, triangles) proteins at the beginning of the first bound state. Parameter values are $\lambda = 0, \tilde{\lambda} = 10, \mu = 1, \theta = 0.5, \psi = 0.25$

In this example ψ is relatively high, this causes a relatively short time in the unbound state and hence a relatively small probability of extinction.

Figure 3 displays the distribution of the number of proteins, in both states. This distribution was computed using the presented algorithm starting at position $n = 30$, as starting at $n = 1$ led to numerical problems. As $\tilde{\lambda} > \lambda$ the feedback is positive. Note that the distribution of proteins in the unbound state shows two maxima. The bimodality of the distribution of the sum of proteins in both states was already demonstrated and discussed elsewhere (Hornos et al., 2005; Ramos et al., 2011), albeit in this example it is shown solely for the unbound state. The unbound state is relatively short, so its protein distribution depends strongly on the protein distribution of the bound state. The left maximum is caused mainly by proteins at the end of the unbound state, whereas the right maximum is caused by proteins at the beginning of the unbound state.

Different extensions of the model are possible. So the gene can have more than two expression states and the produced protein may have to be further metabolized for example by dimerization to bind to the promoter. As mass-action kinetics are assumed for each reaction in the model, not all important features of gene expression may be captured.

In Section 2.3 the influence of feedback on the system is considered. Given that a gene is regulated by its product, and this product is only produced by this gene, the features of feedback compared to an equivalent situation, in which proteins are produced at the same rate but the state of the gene is switched independent of the protein number, is considered.

In figure 4 both situations are compared. It is visible, that in the NFM the

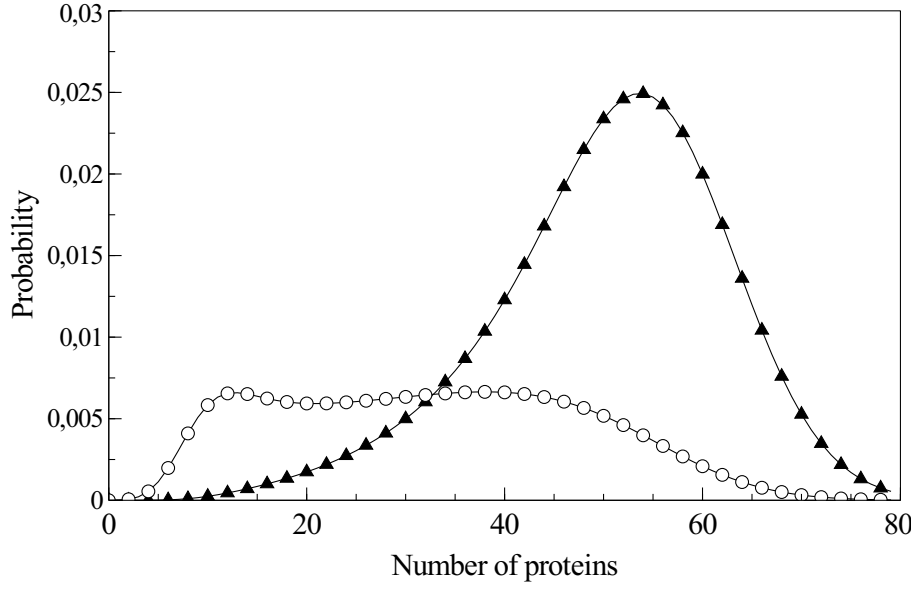


Figure 3: Distribution of (N, A) (circles for $A = 1$ and triangles for $A = 0$). Parameter values are $\lambda = 6, \tilde{\lambda} = 60, \mu = 1, \theta = 1, \psi = 0.066$. Only even protein numbers are displayed.

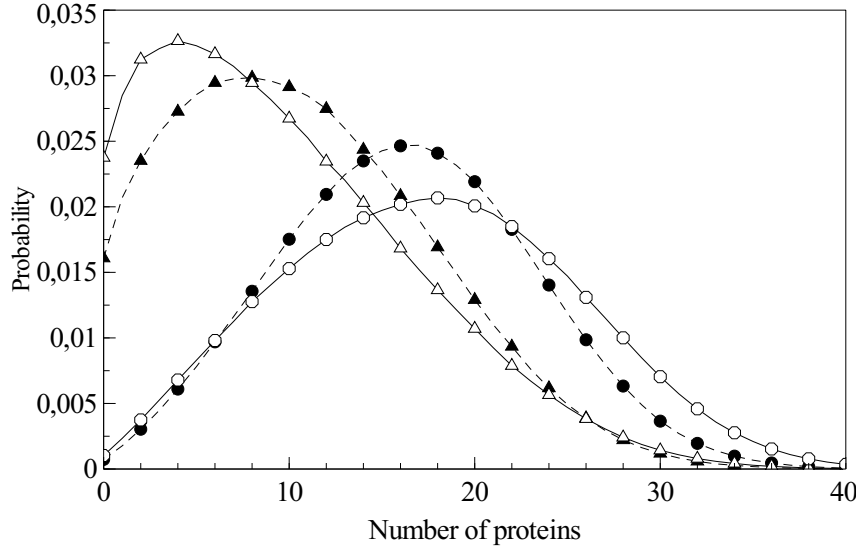


Figure 4: Steady state distribution of the MFM (full markers, dashed line) and the NFM (hollow markers). Circles mark the unbound state and triangles mark the bound state. Only even protein numbers are displayed. Parameter values are $\lambda = 30, \tilde{\lambda} = 0, \mu = 1, \theta = 1.3$ and $\psi = 0.1$. $\tilde{\psi}$ is chosen as described in (2.37), hence $P\{\tilde{A} = 1\} = P\{\bar{A} = 1\}$.

total protein variance and the expected number proteins in the unbound state are higher.

For the MFM and the NFM algorithms similar to the one described in Section 2.2 can be used to determine the distributions of \bar{N} and M respectively.

Given the same total protein production rate, and assuming that besides gene expression there is no influx of protein to the cytoplasm, the models suggest that non-feedback leads to more (less) distinguishable distributions of protein numbers in the two states, if $\lambda > \tilde{\lambda}$ ($\lambda < \tilde{\lambda}$). So, if the protein concentration should be low in the unbound and high in the bound state, non-feedback is superior in this framework.

3.1 Method of Hornos

Previous work (Hornos et al., 2005; Ramos et al., 2011) give an exact solution for the protein distribution in the steady state in this model framework. This is done by rewriting the master equation as partial differential equation, which can be totally integrated using the confluent Heun function. The probability $P\{X = (n, j)\}$, with $n \geq 2$ and $j \in \{0, 1\}$, can be computed. For

$$\begin{aligned} a &:= 1 + \frac{\theta}{\mu + \psi} \left(1 + \frac{\psi\lambda}{\mu\lambda - (\mu + \psi)\tilde{\lambda}} \right), \\ b &:= 1 + \frac{\theta}{\mu + \psi} + \frac{\psi\lambda}{(\mu + \psi)^2}, \\ \eta(z) &:= -\frac{\left[\tilde{\lambda} \left(1 + \frac{\psi}{\mu} \right) - \lambda \right] [(\mu + \psi)z - \mu]}{(\mu + \psi)^2} \end{aligned}$$

it holds

$$P\{X = (n, 1)\} = \frac{A}{n!} \sum_{s=0}^n \binom{n}{s} \left(\frac{\tilde{\lambda}}{\mu} \right)^{n-s} \left(\frac{d\eta}{dz} \right)^s \frac{(a)_s}{(b)_s} M(a + s, b + s, \eta(0)),$$

where $(\cdot)_n$ and M denote the rising factorial and the Kummer function (Abramowitz and Stegun, 1970). A is a normalization constant, guaranteeing, that the sum of the probabilities is 1.

So one has to evaluate n Kummer functions. However evaluating the Kummer function is numerical sophisticated (Muller, 2001). In contrast the recursion method is easier to apply and more intuitive.

Comparing the recursion method with the method of Hornos in numerous simulations, it can be concluded, that both have their advantages and disadvantages. Partly they can complement each other. So the method of Hornos works well for small $\frac{\lambda + \tilde{\lambda}}{2\mu}$ but a wide range of θ and ψ , due to the numerical evaluation of the Kummer function, whereas the recursion method works well for a wide range of $\frac{\lambda + \tilde{\lambda}}{2\mu}$, but only for a relatively small range of θ and ψ . The recursion method has the disadvantage, that minimal changes of probabilities, which are close to machine precision, might lead to enormous changes in the outcome of the recursion. Figure 5 demonstrates the areas, in which both algorithms work well for two examples. In this context an algorithm works well, if the resulting distribution is non negative and asymptotic to zero for large numbers of proteins. The distribution of protein-numbers is evaluated between 0 and the

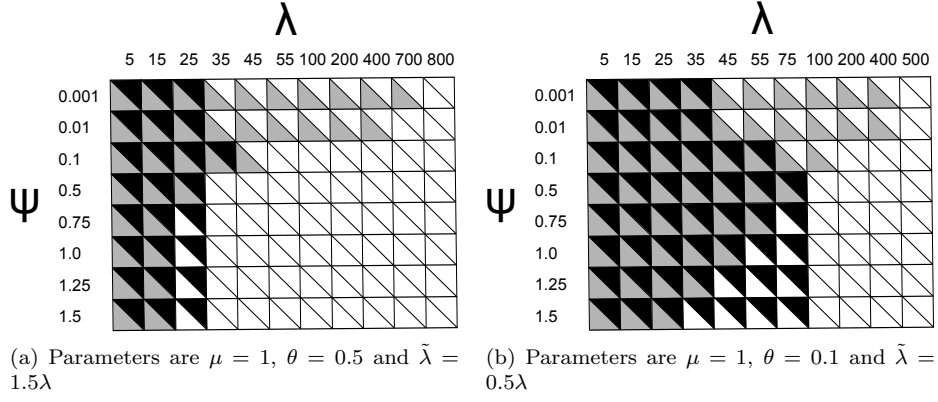


Figure 5: Comparism of both algorithms. An upper black triangle indicates, that the algorithm proposed by Hornos works well, a lower grey triangle indicates, that the algorithm described in Section 2.2 works well

triple of the smallest integer, below which the majorant distribution, which is Poisson, has at least 99.9% of its mass. Hence this integer is the smallest $n \in \mathbb{N}$, for which holds

$$e^{-\frac{\max(\lambda, 1+\tilde{\lambda})}{\mu}} \sum_{k=0}^n \frac{\max(\lambda, 1+\tilde{\lambda})^k}{\mu^k k!} \geq 0.999.$$

4 Proofs

4.1 Proof Theorem 1

The equation (2.2) can be derived by the master equation in the steady state inductively.

The equation (2.3) can be derived by summation over all n of the master equation in the steady state for $A = 1$ or $A = 0$. The equations (2.4) and (2.5) are direct consequences of the master equation evaluated in the steady state.

4.2 Proof Corollary 2.1

Summation over n of (2.2) yields to the equation (2.6). The equation (2.7) follows directly from (2.3). The equation (2.8) follows directly from (2.6) and (2.7). The recursions (2.8), (2.11) and (2.12) are direct consequences from (2.2), (2.4) and (2.5). (2.10) follow directly from (2.6).

4.3 Proof Theorem 2

For the proof the following lemma is needed:

Lemma 4.1. *Let $s \in \mathbb{N}$ and Y be a positive discrete random variable with $\mathbb{E}[Y^s] < \infty$, it holds that*

$$\sum_{i=1}^{\infty} i^s(i+1)P\{Y=i+1\} = \sum_{j=0}^s \binom{s}{j} (-1)^j \mathbb{E}[Y^{s-j+1}].$$

Proof. If $\mathbb{E}[Y^s] < \infty$, it follows that

$$\begin{aligned} \sum_{i=1}^{\infty} i^s(i+1)P\{Y=i+1\} &= \sum_{i=2}^{\infty} (i-1)^s i P\{Y=i\} \\ &= \sum_{i=2}^{\infty} \sum_{j=0}^s \binom{s}{j} i^{s-j+1} (-1)^j P\{Y=i\} \\ &= \sum_{j=0}^s \binom{s}{j} (-1)^j \left[\sum_{i=0}^{\infty} i^{s-j+1} P\{Y=i\} - P\{Y=1\} \right] \\ &= \sum_{j=0}^s \binom{s}{j} (-1)^j [\mathbb{E}[Y^{s-j+1}] - P\{Y=1\}] = \sum_{j=0}^s \binom{s}{j} (-1)^j \mathbb{E}[Y^{s-j+1}]. \end{aligned}$$

□

For the sake of brevity set $\pi_{(n,i)} := P\{X=(n,i)\}$ in this proof for $n \in$

$\mathbb{N}, i \in \{0, 1\}$. It holds that

$$\begin{aligned}
\mathbb{E}[1_{A=1}N^s] &= \sum_{i=0}^{\infty} i^s \pi_{(i,1)} \\
&\stackrel{(2.2)}{=} \frac{1}{\lambda} \left[\sum_{i=1}^{\infty} i^s \left(i\mu\pi_{(i+1,1)} + i\mu\pi_{(i+1,0)} + \mu\pi_{(i+1,1)} - \tilde{\lambda}\pi_{(i,0)} \right) \right] \\
&= \frac{\mu}{\lambda} \sum_{i=1}^{\infty} (i-1)^s \left((i-1) (\pi_{(i,1)} + \pi_{(i,0)}) + \pi_{(i,1)} \right) - \frac{\tilde{\lambda}}{\lambda} \mathbb{E}[N^s 1_{A=0}] \\
&= \frac{\mu}{\lambda} \left(\sum_{j=0}^s \binom{s}{j} (-1)^j (\mathbb{E}[N^{s-j+1}] - \mathbb{E}[N^{s-j} 1_{A=0}]) \right) - \frac{\tilde{\lambda}}{\lambda} \mathbb{E}[N^s 1_{A=0}].
\end{aligned}$$

It follows for the $(s+1)$ -th moment, that

$$\begin{aligned}
\mathbb{E}[N^{s+1}] &= \frac{\lambda}{\mu} \mathbb{E}[N^s 1_{A=1}] + \mathbb{E}[N^s 1_{A=0}] + \frac{\tilde{\lambda}}{\mu} \mathbb{E}[N^s 1_{A=0}] \\
&\quad - \sum_{j=1}^s \binom{s}{j} (-1)^j (\mathbb{E}[N^{s-j+1}] - \mathbb{E}[N^{s-j} 1_{A=0}]).
\end{aligned}$$

Next $\mathbb{E}[1_{A=1}N^s]$ is derived. It holds

$$\begin{aligned}
\mathbb{E}[1_{A=0}N^s] &= \sum_{i=0}^{\infty} i^s \pi_{(i,0)} = \sum_{i=1}^{\infty} i^s \pi_{(i,0)} \\
&\stackrel{(2.4)}{=} \frac{1}{\theta} \left[\sum_{i=1}^{\infty} i^{s+1} (\mu + \psi) \pi_{(i,1)} + i^s \pi_{(i,1)} \lambda - \mu i^s (i+1) \pi_{(i+1,1)} - \lambda i^s \pi_{(i-1,1)} \right] \\
&= \frac{1}{\theta} \left[\mathbb{E}[1_{A=1}N^{s+1}] (\mu + \psi) + \lambda \mathbb{E}[1_{A=1}N^s] \right. \\
&\quad \left. - \sum_{i=1}^{\infty} [\mu i^s (i+1) \pi_{(i+1,1)} + \lambda i^s \pi_{(i-1,1)}] \right] \\
&\stackrel{(\text{Lemma 4.1})}{=} \frac{1}{\theta} \left[\mathbb{E}[1_{A=1}N^{s+1}] \psi \right. \\
&\quad \left. - \sum_{j=1}^s \binom{s}{j} ((-1)^j \mu \mathbb{E}[1_{A=1}N^{s-j+1}] + \lambda \mathbb{E}[1_{A=1}N^{s-j}]) \right].
\end{aligned}$$

Hence it follows

$$\begin{aligned}
&\mathbb{E}[1_{A=1}N^{s+1}] \\
&= \frac{1}{\psi} \left[\theta \mathbb{E}[1_{A=0}N^s] + \sum_{j=1}^s \binom{s}{j} ((-1)^j \mu \mathbb{E}[1_{A=1}N^{s-j+1}] + \lambda \mathbb{E}[1_{A=1}N^{s-j}]) \right].
\end{aligned}$$

4.4 Proof Theorem 3

It is straightforward to prove, that the steady state distribution of the Markov chain exists and that it is unique. The recursion is a direct consequence of the

master equation in the steady state and the master equation in the steady state can be derived by the recursions. Hence there is a one-to-one correspondence between the recursion and the corresponding Markov chain. So due to the linearity of the recursion, if a sequence can be derived using the recursions, for which the corresponding sequence of partial sums is bounded, it is proportional to the unique steady state solution.

4.5 Proof Corollary 2.3

As the bound states expected duration is θ and the unbound states expected duration is $\mathbb{E}[N|A=1]\psi$ it holds

$$P\{A=1\} = \frac{\theta}{\theta + \mathbb{E}[N|A=1]\psi}.$$

If $\tilde{\lambda}$ equals λ , the distribution of proteins during the bound state, and the distribution of proteins minus one during the bond state are Poisson with parameter $\frac{\lambda}{\mu}$. It holds that, if the distribution of a random variable Y satisfy the condition

$$P\{Y=n\} = \frac{\lambda}{\mu n} P\{Y=n-1\}, \quad (4.1)$$

Y is Poisson distributed with parameter $\frac{\lambda}{\mu}$. It can be shown that the recursions (2.11) and (2.12) satisfy for $\lambda = \tilde{\lambda}$ the conditions $P\{X=(n,1)\} = \frac{\lambda}{\mu n} P\{X=(n-1,1)\}$ and $P\{X=(n,0)\} = \frac{\lambda}{\mu(n-1)} P\{X=(n-1,0)\}$. Hence they are Poisson as described above and by Theorem 3 they are unique solutions. As consequence it follows, that

$$\mathbb{E}[N|A=1] = \mathbb{E}[N|A=0] - 1 = \frac{\lambda}{\mu}, \text{ hence } P\{A=1\} = \frac{\mu\theta}{\mu\theta + \lambda\psi}.$$

To prove (2.17), all parameters but $\tilde{\lambda}$, which is increased starting from $\tilde{\lambda} = \lambda$, are fixed. It holds for the unbound and the bound state, that the higher the protein, with which the state starts, the higher the expected protein number at the end of the state. For the bound state this is easy to verify, with Theorems 4 and 5. In the unbound state the number of proteins during the state is given as the sum of the two independent processes. There are proteins, which existed at the beginning of the state and decay during the state, and proteins, which are produced in the state and yet not decayed. Hence, if the impact of the number of proteins $n \in \mathbb{N}$ at time 0 of the state on the expected number of proteins at the end of the state is analyzed, only the first process, which is binomial distributed after time t with rate $\exp(-\mu t)$ and n trials, has to be considered. It is straightforward to verify, that the higher n , the shorter the expected time in the unbound state. Furthermore the distribution of the number of proteins at the end of both state, which started with $m \leq n$ proteins is dominated by the corresponding distributions, which started with n proteins. Given that it can be verified, that, if $\tilde{\lambda}$ is increased $\mathbb{E}[N|A=0]$ increases. Using (2.7) and (2.16) the inequality $P\{A=1\} \leq \frac{\mu\theta}{\mu\theta + \max(\lambda, \tilde{\lambda})\psi}$ follows. The lower estimate is proven analogously.

4.6 Proof Theorem 4

If the switching of states is neglected, in both states $A = 1$ and $A = 0$ there are two independent processes regulating the protein-number. At the one hand each protein decays after an exponential distributed waiting time with rate μ . This affects all proteins in the state $A = 1$ and all but one protein in the state $A = 0$. At the other hand there is an exponential distributed production with rate λ and $\tilde{\lambda}$ respectively. Let n be the number of proteins at the beginning of a state. The distribution of the number of non-decayed proteins at time t , given n proteins at time 0 is binomial with rate $\exp(-\mu t)$ and n trials in the state $A = 1$. For the state $A = 0$ there are $n - 1$ trials and the gene-bound protein has to be added to the protein number.

The distribution of the number of produced and not yet decayed proteins at time t is Poisson with rate $\frac{X}{\mu}(1 - \exp(-\mu t))$ with $X = \lambda$ in state $A = 1$ and $X = \tilde{\lambda}$ in state $A = 0$. The bound state duration is distributed exponentially with parameter θ . Hence (starting at $t = 0$) $\exp(-t\theta)\theta$ is the density of the duration of the bound state. Let $Q_{n,t}$ be a random variable describing the number of proteins at time t in the bound state, given that n proteins existed at the beginning of the bound state at time 0 and that the bound state is yet not stopped. It follows, that

$$\mathbb{E}[Q_{n,t}] = (n-1)e^{-t\mu} + 1 + \frac{\tilde{\lambda}}{\mu}(1 - e^{-t\mu}).$$

Hence (2.18) holds, as

$$\begin{aligned} \mathbb{E}[C_1|B_0 = n] &= \mathbb{E}\left[\int_0^\infty Q_{n,t}e^{-t\theta}\theta dt\right] \\ &= \int_0^\infty \mathbb{E}[Q_{n,t}]e^{-t\theta}\theta dt = \left(n - \frac{\tilde{\lambda}}{\mu} - 1\right)\frac{\theta}{\theta + \mu} + 1 + \frac{\tilde{\lambda}}{\mu} = \frac{n\theta + \tilde{\lambda} + \mu}{\theta + \mu}. \end{aligned}$$

The remaining propositions are direct consequences of these features and the features of the binomial- and Poisson-distribution.

4.7 Proof Theorem 5

As the duration of the bound state is exponential distributed, it holds that at each time point in the bound state the propensity of switching to the unbound state is the same. Hence the expected protein number during a bound state equals the expected protein number at the end of a bound state.

4.8 Proof Theorem 6

The first equation follows directly from (2.19).

Using the notation from the proof of Theorem 4, it holds

$$\mathbb{E}[C^2] = \mathbb{E}\left[\frac{\int_0^\infty \exp(-\theta t)\theta \int_0^t Q_{n,s}^2 ds dt}{\int_0^\infty t \exp(-t\theta)\theta dt}\right] = \theta^2 \int_0^\infty \exp(-t\theta) \int_0^t \mathbb{E}[Q_{n,s}^2] ds dt.$$

By the proof of Theorem 4 it holds for $n \geq 1$

$$Q_{n,t} = X_{n-1}(t) + 1 + Y(t).$$

$X_n(t)$ is binomial distributed with rate $\exp(-\mu t)$ and n trials and $Y(t)$ is Poisson distributed with rate $\frac{\tilde{\lambda}}{\mu}(1 - \exp(-\mu t))$. Using the features of the binomial and Poisson distribution it follows for $n \in \mathbb{N}$

$$\begin{aligned}
& \mathbb{E}[Q_{n,t}^2] \\
&= \mathbb{E}[(1 + X_{n-1}(t))^2] + 2\mathbb{E}[(1 + X_{n-1}(t))Y(t)] + \mathbb{E}[Y(t)^2] \\
&= 1 + 2\mathbb{E}[X_{n-1}(t)] + \mathbb{E}[X_{n-1}(t)^2] + 2\mathbb{E}[Y(t)] + 2\mathbb{E}[X_{n-1}(t)Y(t)] + \mathbb{E}[Y(t)^2] \\
&= 1 + 2(n-1)e^{-\mu t} + (n-1)e^{-\mu t}(1 - e^{-\mu t}) + ((n-1)e^{-\mu t})^2 + 2\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t}) \\
&\quad + \frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t}) + \left(\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t})\right)^2 + 2\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t})(n-1)e^{-\mu t} \\
&= 1 + 3(n-1)e^{-\mu t} - (n-1)e^{-2\mu t} + ((n-1)e^{-\mu t})^2 + 3\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t}) \\
&\quad + \left(\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t})\right)^2 + 2\frac{\tilde{\lambda}}{\mu}(1 - e^{-\mu t})(n-1)e^{-\mu t} \\
&= n^2e^{-2\mu t} + n\left(3 + \frac{2\tilde{\lambda}}{\mu}\right)(e^{-\mu t} - e^{-2\mu t}) + 1 + \frac{3\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2 \\
&\quad - e^{-\mu t}\left(3 + \frac{5\tilde{\lambda}}{\mu} + 2\left(\frac{\tilde{\lambda}}{\mu}\right)^2\right) + e^{-2\mu t}\left(2 + \frac{2\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2\right).
\end{aligned}$$

Given that it can be computed

$$\begin{aligned}
& \int_0^\infty e^{-\theta t} \int_0^t \mathbb{E}[Q_{n,s}^2] ds dt = \\
& \int_0^\infty e^{-\theta t} \int_0^t \left(n^2e^{-2s\mu} + n\left(3 + \frac{2\tilde{\lambda}}{\mu}\right)(e^{-s\mu} - e^{-2s\mu}) + 1 + \frac{3\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2 \right. \\
& \quad \left. - e^{-\mu s}\left(3 + \frac{5\tilde{\lambda}}{\mu} + 2\left(\frac{\tilde{\lambda}}{\mu}\right)^2\right) + e^{-2\mu s}\left(2 + \frac{2\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2\right) \right) ds dt \\
&= n^2 \frac{1}{\theta(\theta + 2\mu)} + n\left(3 + \frac{2\tilde{\lambda}}{\mu}\right) \left(\frac{1}{\theta(\theta + \mu)} - \frac{1}{\theta(\theta + 2\mu)} \right) + 1 + \frac{3\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2 \\
&\quad - \frac{1}{\theta(\theta + \mu)} \left(3 + \frac{5\tilde{\lambda}}{\mu} + 2\left(\frac{\tilde{\lambda}}{\mu}\right)^2 \right) + \frac{1}{\theta(\theta + 2\mu)} \left(2 + \frac{2\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu}\right)^2 \right).
\end{aligned}$$

Hence it holds that

$$\begin{aligned}\mathbb{E}[C^2] &= \mathbb{E}[B] \left(3 + \frac{2\tilde{\lambda}}{\mu} \right) \left(\frac{\theta\mu}{(\theta+\mu)(\theta+2\mu)} \right) + \theta^2 \left(1 + \frac{3\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) \\ &\quad - \frac{\theta}{(\theta+\mu)} \left(3 + \frac{5\tilde{\lambda}}{\mu} + 2 \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) + \frac{\theta}{(\theta+2\mu)} \left(2 + \frac{2\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) + \frac{\theta\mathbb{E}[B^2]}{(\theta+2\mu)},\end{aligned}$$

so as consequence it holds

$$\begin{aligned}\mathbb{E}[B^2] &= \frac{\theta+2\mu}{\theta+\mu} \left(3 + 5\frac{\tilde{\lambda}}{\mu} + 2 \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) - \left(2 + 2\frac{\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right) \\ &\quad + \mathbb{E}[C^2] \frac{\theta+2\mu}{\theta} - \frac{\mu\mathbb{E}[B]}{\theta+\mu} \left(3\mu\theta + 2\frac{\tilde{\lambda}}{\mu} \right) - (\theta+2\mu)\theta \left(1 + 3\frac{\tilde{\lambda}}{\mu} + \left(\frac{\tilde{\lambda}}{\mu} \right)^2 \right).\end{aligned}$$

4.9 Proof Theorem 7

The probability, that starting with n proteins all proteins decay in an unbound state, before reaching the bound state is $\left(\frac{\mu}{\psi+\mu}\right)^n$, conversely for $1 \leq m \leq n$ the conditional probability for reaching the bound state with m proteins is $P\{B_1 = m | C_0 = n\} = \frac{\psi}{\psi+\mu} \left(\frac{\mu}{\psi+\mu}\right)^{n-m}$.

For $1 \leq m \leq n$ the expected time for the event that n proteins decay to m proteins in the unbound state before it switches to the bound state is $\frac{1}{\mu} \sum_{i=m+1}^n \frac{1}{i} + \frac{1}{m\psi}$. The expected time in the bound state is $\frac{1}{\theta}$. With this the recursions for $P\{S = i | X_0 = (n, 0)\}$ and $\mathbb{E}[T_i | X_0 = (n, 0), i < S]$ and the expectation values $\mathbb{E}[S | X_0 = (n, 0)]$ and $\mathbb{E}[S^2 | X_0 = (n, 0)]$ can be derived.

To prove the last claim, features of the geometric series are used. It holds for $n \geq 1$, that

$$\begin{aligned}\mathbb{E}[B_1 | C_0 = n] &= \sum_{i=0}^n i \frac{\psi}{\psi+\mu} \left(\frac{\mu}{\mu+\psi} \right)^i \\ &= \frac{(\mu+\psi)^2}{\psi\mu} \left[n \left(\frac{\mu}{\mu+\psi} \right)^{n+2} - (n+1) \left(\frac{\mu}{\mu+\psi} \right)^{n+1} + \frac{\mu}{\mu+\psi} \right] \\ &= \frac{\mu+\psi}{\psi} \left(1 - \left(\frac{\mu}{\mu+\psi} \right)^n \left(1 + \frac{n\psi}{\mu+\psi} \right) \right).\end{aligned}$$

4.10 Proof Corollary 2.7

$\mathbb{E}[C_2 | C_1 = n]$ can be computed using the expression for $\mathbb{E}[B_1 | C_0 = n]$ from Theorem 7. Let B_0 be the distribution of proteins at the end of the unbound

state, between the bound states ending with C_1 and C_2 . It holds

$$\begin{aligned}
\mathbb{E}[C_2|C_1 = n] &= \sum_{m=1}^{\infty} mP\{C_2 = m|C_1 = n\} \\
&= \sum_{b=1}^n P\{B_0 = b|C_1 = n\} \sum_{m=1}^{\infty} mP\{C_2 = m|B_0 = b\} \\
&\stackrel{(2.18)}{=} \sum_{b=1}^n P\{B_0 = b|C_1 = n\} \left(\frac{b\theta + \tilde{\lambda} + \mu}{\theta + \mu} \right) \\
&= \frac{\theta}{\theta + \mu} \mathbb{E}[B_0|C_1 = n] + \frac{\tilde{\lambda} + \mu}{\theta + \mu}.
\end{aligned}$$

Monotony of $\mathbb{E}[C_2|C_1 = n]$ is straightforward to show. So for $n \geq 1$ it holds

$$\mathbb{E}[C_2|C_1 = 1] \leq \mathbb{E}[C_2|C_1 = n] \leq \lim_{m \rightarrow \infty} \mathbb{E}[C_2|C_1 = m],$$

$$\begin{aligned}
\lim_{m \rightarrow \infty} \mathbb{E}[C_2|C_1 = m] &= \frac{\tilde{\lambda} + \mu}{\theta + \mu} + \frac{\theta}{\theta + \mu} \frac{\mu + \psi}{\psi}, \\
\mathbb{E}[C_2|C_1 = 1] &= \frac{\tilde{\lambda} + \mu}{\theta + \mu} + \frac{\theta}{\theta + \mu} \frac{\psi}{\mu + \psi}.
\end{aligned}$$

The distribution of the totally produced proteins is easy to compute given the number of bound states. During one bound state m proteins are produced with probability $\frac{\theta}{\theta + \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\theta + \tilde{\lambda}} \right)^m$. $\binom{m+i-1}{i}$ is the number of multisets of cardinality i , hence the number of possibilities to express $m \in \mathbb{N}$ as sum of i positive integers including zero. For the distribution after i states, one has to take $\frac{\theta}{\theta + \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\theta + \tilde{\lambda}} \right)^m \binom{m+i-1}{i}$ times.

4.11 Proof Proposition 2.9

Let $z := \frac{\tilde{\psi}}{\psi}$, using (2.34) and (2.37) it can be derived, that $\mathbb{E}[\bar{N}|A = 1] = z$. Consider C , the number of proteins at the beginning of the unbound state. If C is transformed by an unbound state of the NFM, it holds that for all protein numbers smaller than z the hazard of ending the unbound state is greater than in the MFM. As the expected duration of the unbound state in both models is assumed equal (and does not depend on the start position in the NFM) these higher hazard must be compensated by protein numbers greater z , which have a lower hazard of ending the unbound state in the NFM. Hence the expected number of proteins after the unbound state in the NFM is greater than z . Furthermore by (2.32), the expected number of proteins after (and during) an unbound state in the NFM is linear dependent to the expected number of proteins at the beginning of this unbound state. Hence, if $\mathbb{E}[M|\tilde{A} = 0] \geq \mathbb{E}[\bar{N}|\tilde{A} = 0]$, it follows $\mathbb{E}[M|\tilde{A} = 1] > \mathbb{E}[\bar{N}|\tilde{A} = 1]$, which contradicts $\mathbb{E}[\bar{N}] = \mathbb{E}[M]$. Hence the equation (2.38) holds.

(2.39) follows directly by assuming (2.37) and using (2.30), (2.36) and the fact

that there is a $\epsilon \geq 0$, so that

$$\begin{aligned}\mathbb{E}[1_{\tilde{A}=1}M] &= \mathbb{E}[1_{\tilde{A}=1}\tilde{N}] + \epsilon, \\ \mathbb{E}[1_{\tilde{A}=0}M] &= \mathbb{E}[1_{\tilde{A}=0}\tilde{N}] - \epsilon\end{aligned}$$

holds.

4.12 Proof Proposition 2.10

Set $c := \frac{\lambda}{\mu}$ constant. The effect of changing μ (and consequently $\lambda = c\mu$) on the distribution $X = (N, A)$ is considered. If μ tends to ∞ , $\mathbb{E}[N|A=1]$ tends to c . Thus it holds

$$\lim_{\mu \rightarrow \infty} P\{A=1\} = \frac{\theta}{\theta + c\psi} \quad (4.2)$$

This is by (2.17) the lowest possible value, which $P\{A=1\}$ can attain. Using (2.4) it can be seen, that for all $n \in \mathbb{N}$

$$\lim_{\mu \rightarrow 0} P\{X=(n,0)\} = \frac{\psi n}{\theta} P\{X=(n,1)\}$$

Hence as $P\{X=(0,0)\} = 0$, it always holds $\mathbb{E}[N|A=1] < \mathbb{E}[N] < \mathbb{E}[N|A=0]$.

Using (2.7) it follows

$$\lim_{\mu \rightarrow 0} P\{A=1\} > \frac{\theta}{\theta + \mathbb{E}[N]\psi} \geq \frac{\theta}{\theta + c\psi} \quad (4.3)$$

Let $\tilde{\lambda} = 0$, set

$$E_0 := \frac{1}{2\psi} \left(\psi - \theta + \sqrt{(\psi - \theta)^2 + 4\theta\psi c} \right) = \mathbb{E}[\tilde{N}].$$

Hence E_0 is constant, if the ratio c and θ, ψ are fixed. If μ tends to zero, it can be derived by using (2.6) and (4.3), that

$$\lim_{\mu \rightarrow 0} \mathbb{E}[N] \geq E_0.$$

Furthermore using (4.2) and (2.6) it holds

$$\lim_{\mu \rightarrow \infty} \mathbb{E}[N] = E_\infty := c \frac{\theta + \psi}{\theta + c\psi}$$

Hence $E_0 - E_\infty$ is greater equal the maximal distance between both approaches.

Acknowledgements

The author would like to thank Prof. Peter Pfaffelhuber from the department of statistics, University of Freiburg, for making helpful comments on the manuscript.

References

- Abramowitz, M. and I. A. Stegun (1970). Handbook of mathematical function with formulas, graphs, and mathematical tables. *National Bureau of Standards, Applied Mathematics Series 55*.
- Ackers, G. K., A. D. Johnson, and M. A. Shea (1982). Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences* 79(4), 1129–1133.
- Anderson, D. F. and T. G. Kurtz (2011). Continuous time markov chain models for chemical reaction networks. In *Design and Analysis of Biomolecular Circuits*, pp. 3–42. Springer.
- Ay, A. and D. N. Arnosti (2011). Mathematical modeling of gene expression: a guide for the perplexed biologist. *Critical reviews in biochemistry and molecular biology* 46(2), 137–151.
- De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology* 9(1), 67–103.
- Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002). Stochastic gene expression in a single cell. *Science Signaling* 297(5584), 1183.
- Gunawardena, J. (2003). Chemical reaction network theory for in-silico biologists. *Notes available for download at <http://vcp.med.harvard.edu/papers/crnt.pdf>*.
- Hornos, J., D. Schultz, G. Innocentini, J. Wang, A. Walczak, J. Onuchic, and P. Wolynes (2005). Self-regulating gene: An exact solution. *Physical Review E* 72(5), 051907.
- Jacob, F. and J. Monod (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* 3(3), 318–356.
- McAdams, H. H. and A. Arkin (1999). It’s a noisy business! genetic regulation at the nanomolar scale. *Trends in Genetics* 15(2), 65–69.
- Muller, K. E. (2001). Computing the confluent hypergeometric function, $m(a, b, x)$. *Numerische Mathematik* 90(1), 179–196.
- Paulsson, J. (2005). Models of stochastic gene expression. *Physics of life reviews* 2(2), 157–175.
- Peccoud, J. and B. Ycart (1995). Markovian modeling of gene-product synthesis. *Theoretical population biology* 48(2), 222–234.
- Ptashne, M. (1992). *A Genetic Switch: Phage Lambda and Higher Organisms. 2nd edition*. Cambridge, MA: Cell Press and Blackwell Science Publications.
- Ramos, A., G. Innocentini, and J. Hornos (2011). Exact time-dependent solutions for a self-regulating gene. *Physical Review E* 83(6), 062902.
- Zhu, R. and D. Salathé (2008). Delay stochastic simulation of single-gene expression reveals a detailed relationship between protein noise and mean abundance. *FEBS letters* 582(19), 2905–2910.